

# Data visualisation is important

Visualisations aren't just furniture for publications



@martijnhadley

<http://www.tradingplacesproperty.com/content/post/what-is-fair-wear-and-tear>

*“Most scientific data is created in a form and organization that facilitates its generation rather than focusing on its eventual use.”*

**Table 2.** Successes and failures for articles with non-zero metric scores, aggregated by journal, and only including journals for which there is at least one success or failure.

Metric+	Mostly success	Mostly failure	Z	Equal	Journals
Tweets**	1097 (58%)	646 (34%)	10.8	148 (8%)	1891
**	1032 (59%)	586 (33%)	11.1	139 (8%)	1757
FbWalls**	414 (53%)	282 (36%)	5.0	86 (11%)	782
**	308 (55%)	188 (34%)	5.4	62 (11%)	558
RH	276 (51%)	221 (41%)	2.5	47 (9%)	544
	193 (51%)	157 (41%)	1.9	30 (8%)	380
Blogs**	190 (58%)	104 (32%)	5.0	32 (10%)	326
**	129 (57%)	70 (31%)	4.2	26 (12%)	225
Google+	61 (50%)	53 (44%)	0.7	7 (6%)	121
	25 (48%)	24 (46%)	0.1	3 (6%)	52
MSM	29 (56%)	17 (33%)	1.8	6 (12%)	52
	13 (52%)	9 (36%)	0.9	3 (12%)	25
Reddits	22 (51%)	17 (40%)	0.8	4 (9%)	43
	9 (47%)	7 (37%)	0.5	3 (16%)	19
Forums	5 (83%)	1 (17%)	1.6	0 (0%)	6
	3 (100%)	0 (0%)	1.7	0 (0%)	3
Q&A	4 (67%)	1 (17%)	1.3	1 (17%)	6
	2 (67%)	0 (0%)	1.4	1 (33%)	3
Pinner	2 (67%)	1 (33%)	0.6	0 (0%)	3
	0 (—%)	0 (—%)	-	0 (—%)	0
LinkedIn	0 (—%)	0 (—%)	-	0 (—%)	0
	0 (—%)	0 (—%)	-	0 (—%)	0

+ In each cell the upper figure is for all journals and the lower figure is for journals with at least 10 articles tested. \* Ratio of successes to failures significantly different from 0.5 at  $p=0.05$ , \*\* Significant at  $p=0.01$ ; both Bonferroni corrected for  $n=11$ .  
doi:10.1371/journal.pone.0064841.t002

*“Most scientific data is created in a form and organization that facilitates its generation rather than focusing on its eventual use.”*

*“[... data management] has mostly focused on the efficiency of query-based retrieval of the collected data, rather than on data exploration”*

**Table 2.** Successes and failures for articles with non-zero metric scores, aggregated by journal, and only including journals for which there is at least one success or failure.

Metric+	Mostly success	Mostly failure	Z	Equal	Journals
Tweets**	1097 (58%)	646 (34%)	10.8	148 (8%)	1891
**	1032 (59%)	586 (33%)	11.1	139 (8%)	1757
FbWalls**	414 (53%)	282 (36%)	5.0	86 (11%)	782
**	308 (55%)	188 (34%)	5.4	62 (11%)	558
RH	276 (51%)	221 (41%)	2.5	47 (9%)	544
	193 (51%)	157 (41%)	1.9	30 (8%)	380
Blogs**	190 (58%)	104 (32%)	5.0	32 (10%)	326
**	129 (57%)	70 (31%)	4.2	26 (12%)	225
Google+	61 (50%)	53 (44%)	0.7	7 (6%)	121
	25 (48%)	24 (46%)	0.1	3 (6%)	52
MSM	29 (56%)	17 (33%)	1.8	6 (12%)	52
	13 (52%)	9 (36%)	0.9	3 (12%)	25
Reddits	22 (51%)	17 (40%)	0.8	4 (9%)	43
	9 (47%)	7 (37%)	0.5	3 (16%)	19
Forums	5 (83%)	1 (17%)	1.6	0 (0%)	6
	3 (100%)	0 (0%)	1.7	0 (0%)	3
Q&A	4 (67%)	1 (17%)	1.3	1 (17%)	6
	2 (67%)	0 (0%)	1.4	1 (33%)	3
Pinters	2 (67%)	1 (33%)	0.6	0 (0%)	3
	0 (—%)	0 (—%)	-	0 (—%)	0
LinkedIn	0 (—%)	0 (—%)	-	0 (—%)	0
	0 (—%)	0 (—%)	-	0 (—%)	0

+ In each cell the upper figure is for all journals and the lower figure is for journals with at least 10 articles tested. \* Ratio of successes to failures significantly different from 0.5 at  $p=0.05$ , \*\* Significant at  $p=0.01$ ; both Bonferroni corrected for  $n=11$ .  
doi:10.1371/journal.pone.0064841.t002

# Research data is too often born and buried in a table

What can visualisations be useful for?

@martijnhadley

**Table 2.** Successes and failures for articles with non-zero metric scores, aggregated by journal, and only including journals for which there is at least one success or failure.

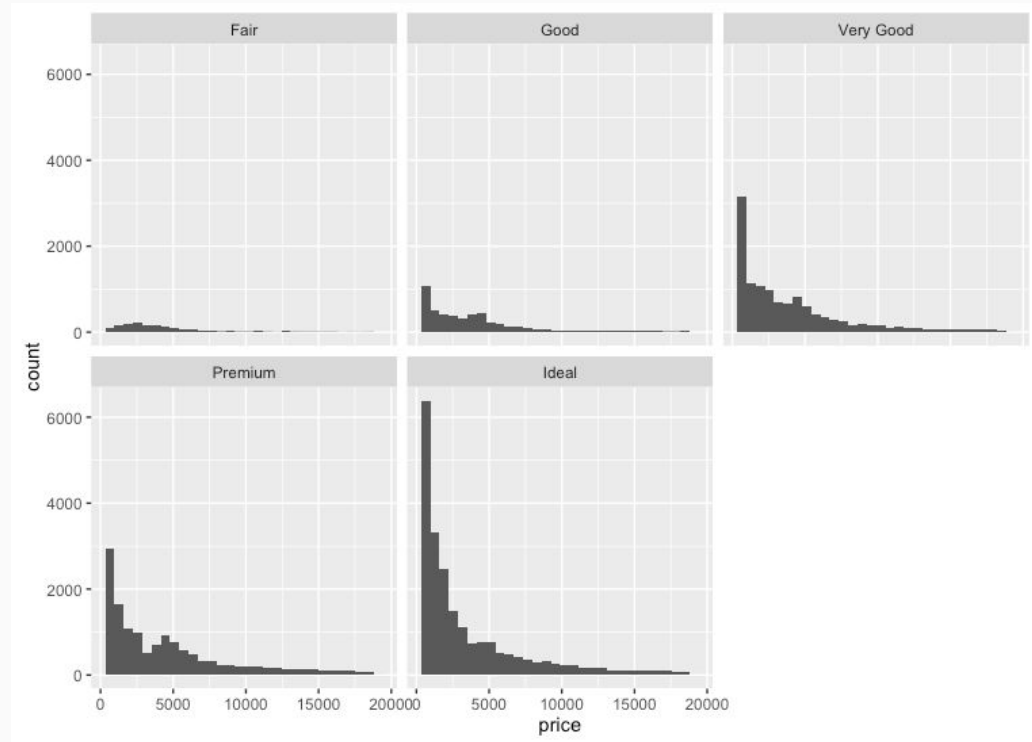
Metric+	Mostly success	Mostly failure	Z	Equal	Journals
Tweets**	1097 (58%)	646 (34%)	10.8	148 (8%)	1891
**	1032 (59%)	586 (33%)	11.1	139 (8%)	1757
FbWalls**	414 (53%)	282 (36%)	5.0	86 (11%)	782
**	308 (55%)	188 (34%)	5.4	62 (11%)	558
RH	276 (51%)	221 (41%)	2.5	47 (9%)	544
	193 (51%)	157 (41%)	1.9	30 (8%)	380
Blogs**	190 (58%)	104 (32%)	5.0	32 (10%)	326
**	129 (57%)	70 (31%)	4.2	26 (12%)	225
Google+	61 (50%)	53 (44%)	0.7	7 (6%)	121
	25 (48%)	24 (46%)	0.1	3 (6%)	52
MSM	29 (56%)	17 (33%)	1.8	6 (12%)	52
	13 (52%)	9 (36%)	0.9	3 (12%)	25
Reddits	22 (51%)	17 (40%)	0.8	4 (9%)	43
	9 (47%)	7 (37%)	0.5	3 (16%)	19
Forums	5 (83%)	1 (17%)	1.6	0 (0%)	6
	3 (100%)	0 (0%)	1.7	0 (0%)	3
Q&A	4 (67%)	1 (17%)	1.3	1 (17%)	6
	2 (67%)	0 (0%)	1.4	1 (33%)	3
Pinners	2 (67%)	1 (33%)	0.6	0 (0%)	3
	0 (-%)	0 (-%)	-	0 (-%)	0
LinkedIn	0 (-%)	0 (-%)	-	0 (-%)	0
	0 (-%)	0 (-%)	-	0 (-%)	0

+ In each cell the upper figure is for all journals and the lower figure is for journals with at least 10 articles tested. \* Ratio of successes to failures significantly different from 0.5 at  $p = 0.05$ , \*\* Significant at  $p = 0.01$ ; both Bonferroni corrected for  $n = 11$ .

doi:10.1371/journal.pone.0064841.t002

# What can visualisations be useful for?

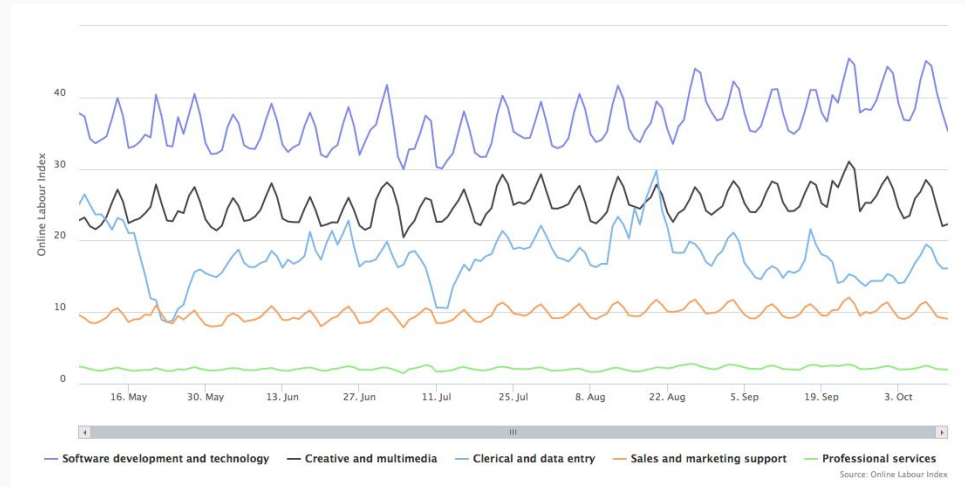
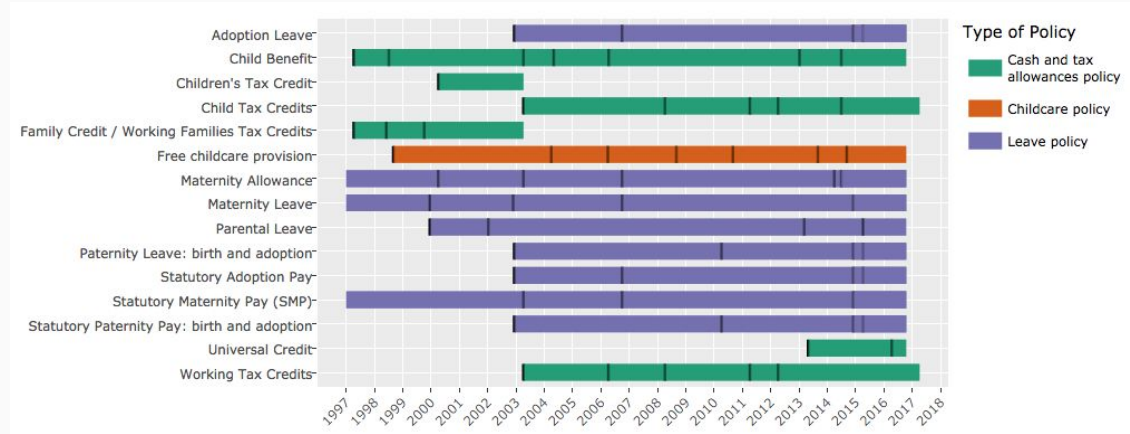
- Exploratory data analysis



You don't know what's interesting until you take a look...

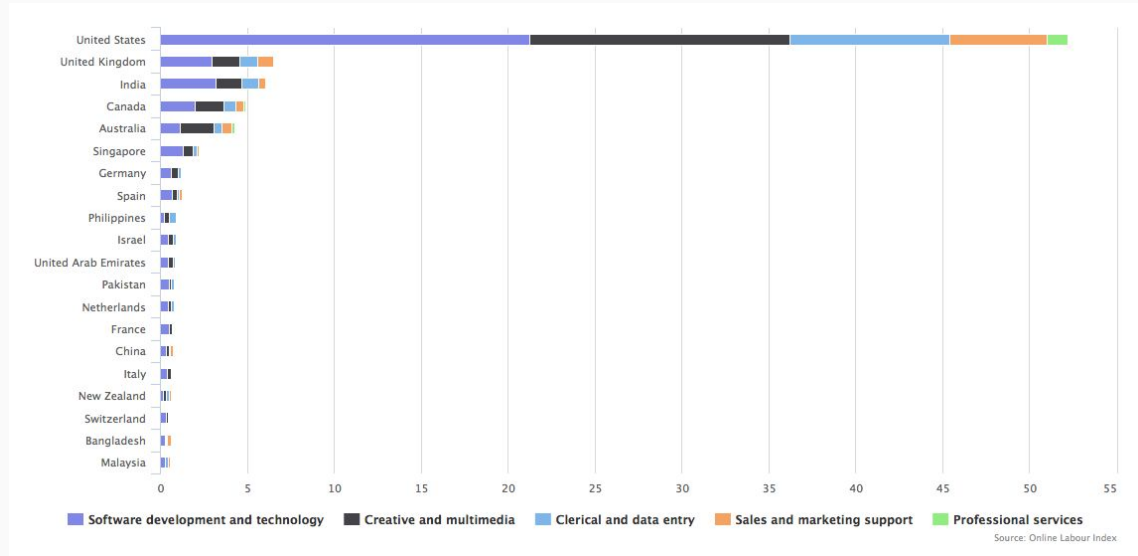
# What can visualisations be useful for?

- Exploratory data analysis
- Summarise trends in an easily consumable manner



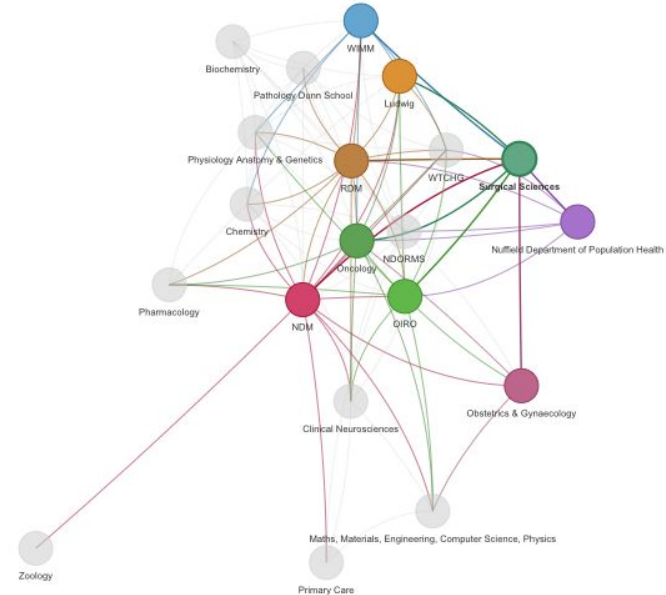
# What can visualisations be useful for?

- Exploratory data analysis
- Summarise trends in an easily consumable manner
- Physically demonstrate comparisons between groups of data



# What can visualisations be useful for?

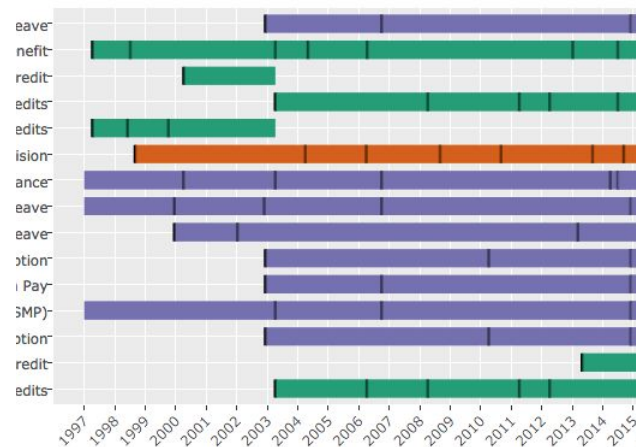
- Exploratory data analysis
- Summarise trends in an easily consumable manner
- Physically demonstrate comparisons between groups of data
- Present connections otherwise difficult to communicate





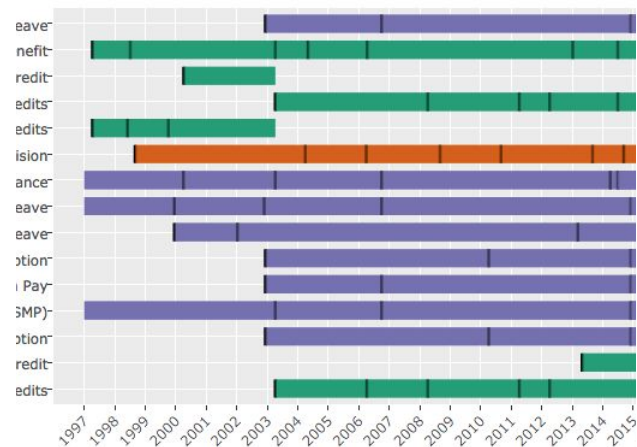
# Why use visualisations?

- Visualisations are easier to parse than long, jargon-filled blocks of text



# Why use visualisations?

- Visualisations are easier to parse than long, jargon-filled blocks of text
- Visualisations can be consumed by the general and expert easily





# Moving beyond dead trees

“... interactivity is the new colour chart...”



@martijnhadley



@photosteve101 (flickr.com/photos/42931449@N07/5263540555/in/album-72157625505293849/)

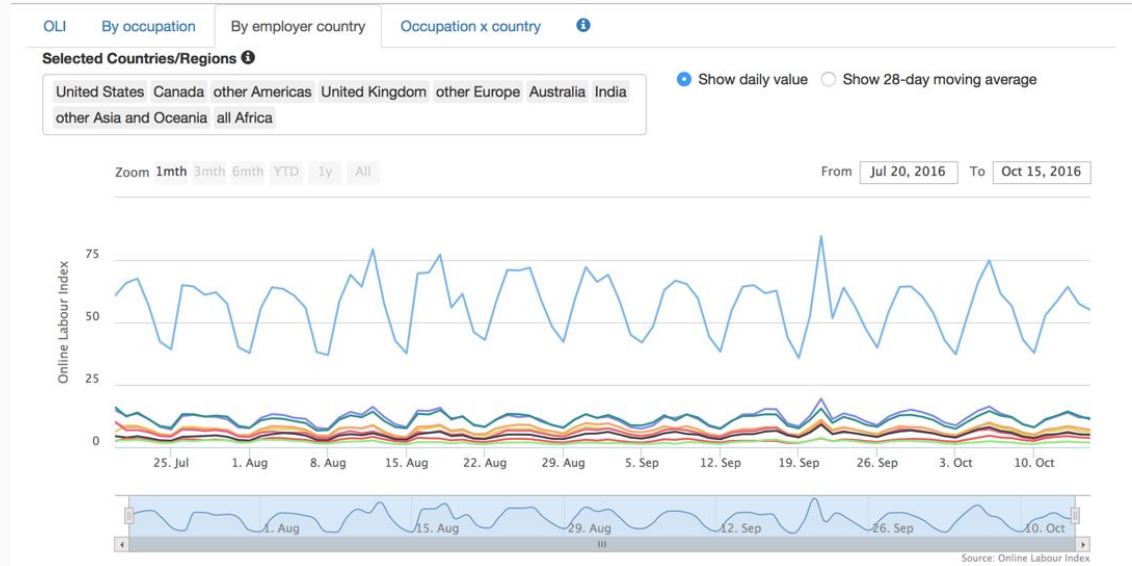
# What does interactivity provide?

- Provide alternative methods to access data



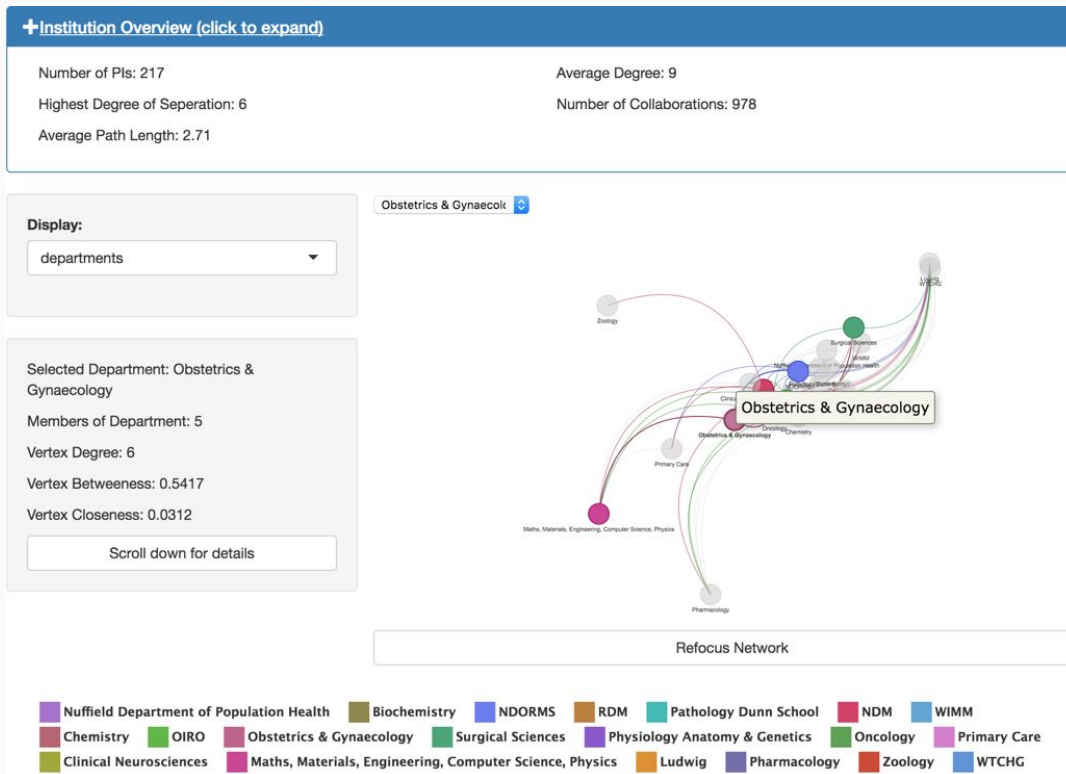
# What does interactivity provide?

- Provide alternative methods to access data
- Allow users to slice through datasets



# What does interactivity provide?

- Provide alternative methods to access data
- Allow users to slice through datasets
- Combine summary and detailed information



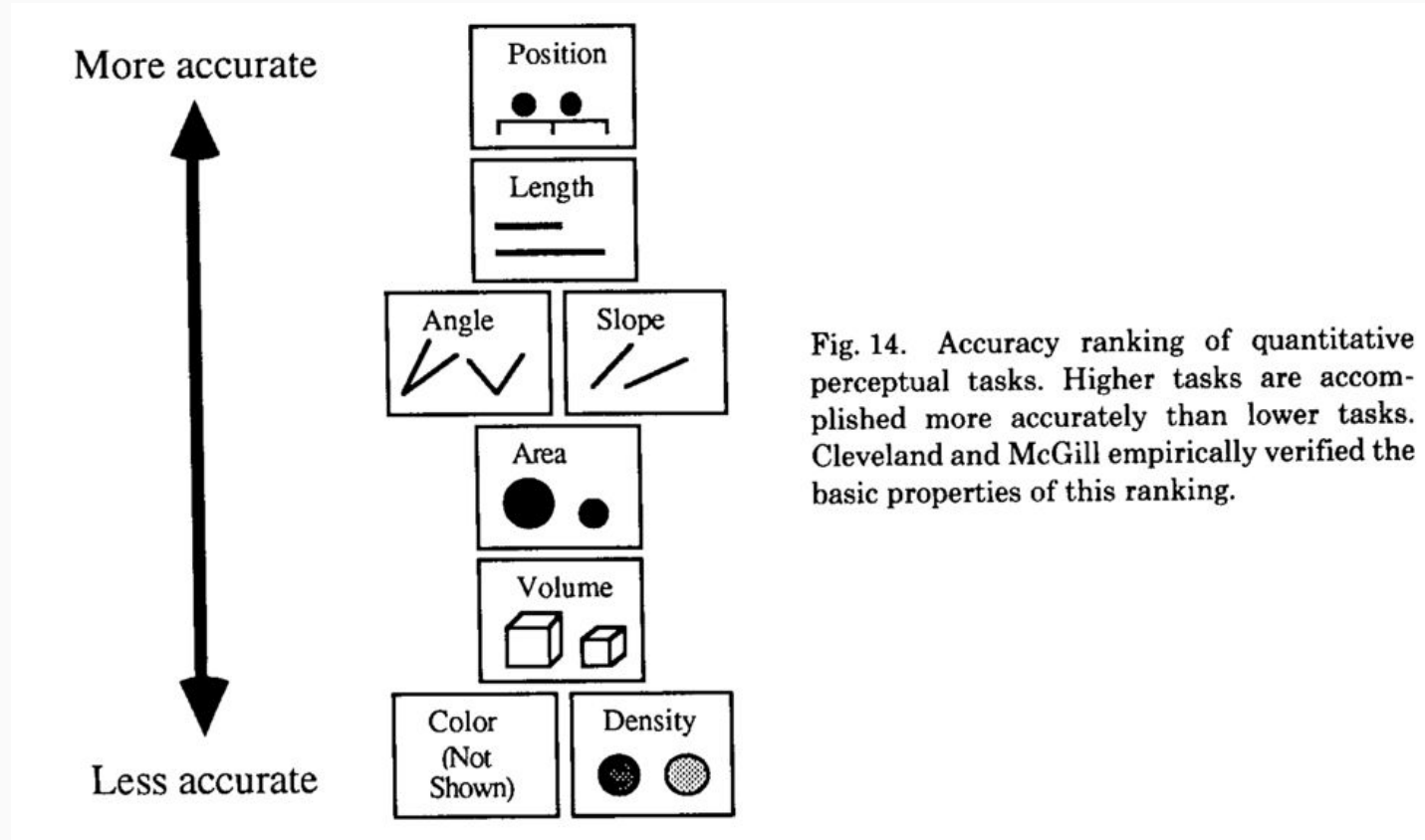
# Designing visualisations to expose data

Putting form over function when communicating data visually

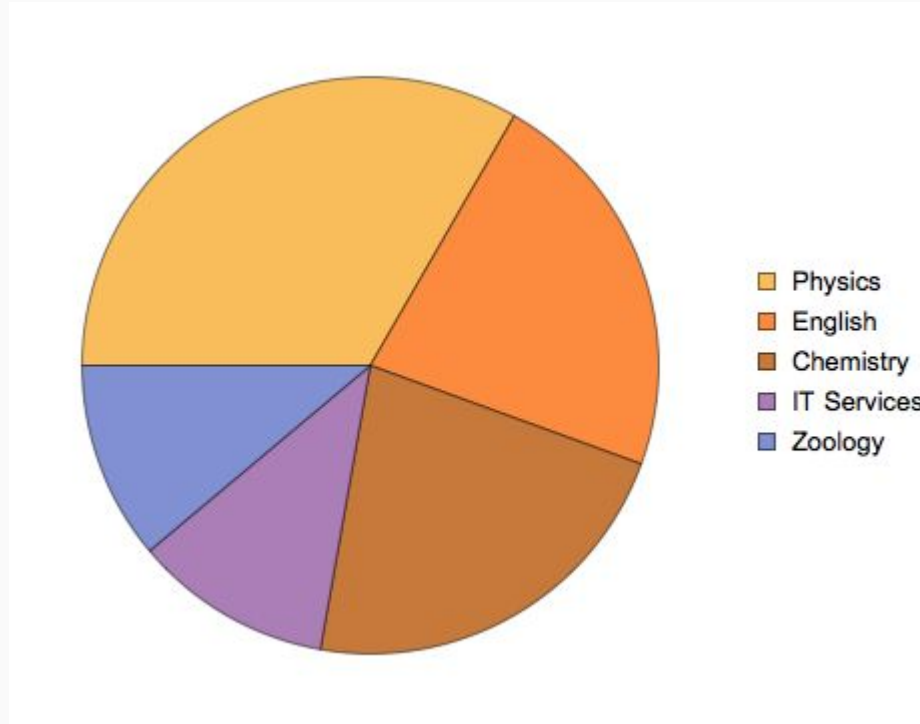




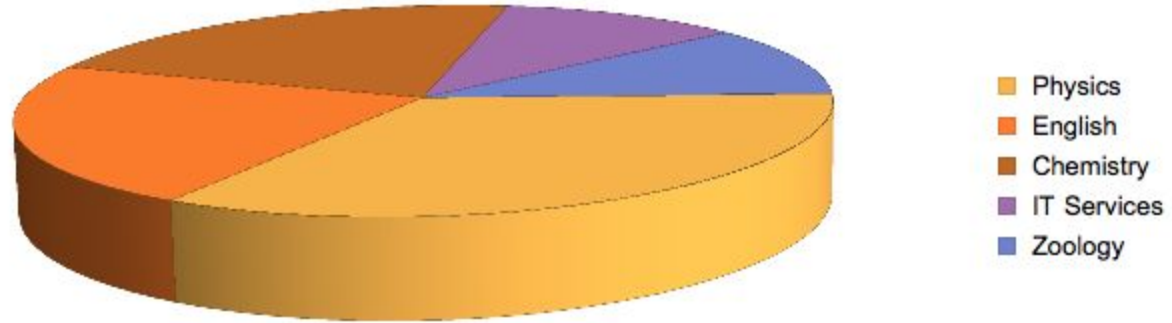
# Graphical Perception Theory



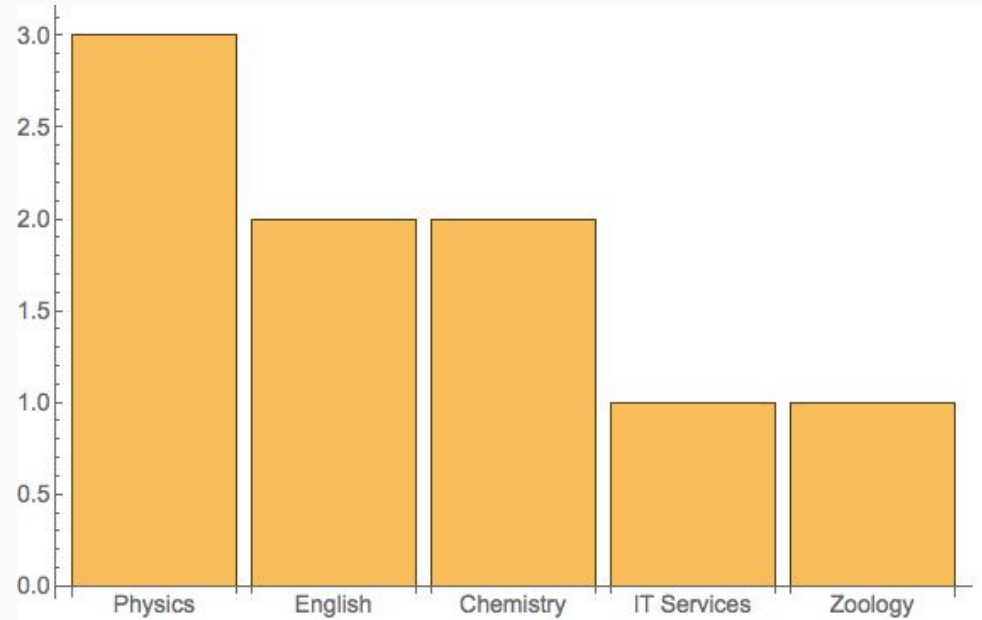
# Graphical Perception Theory



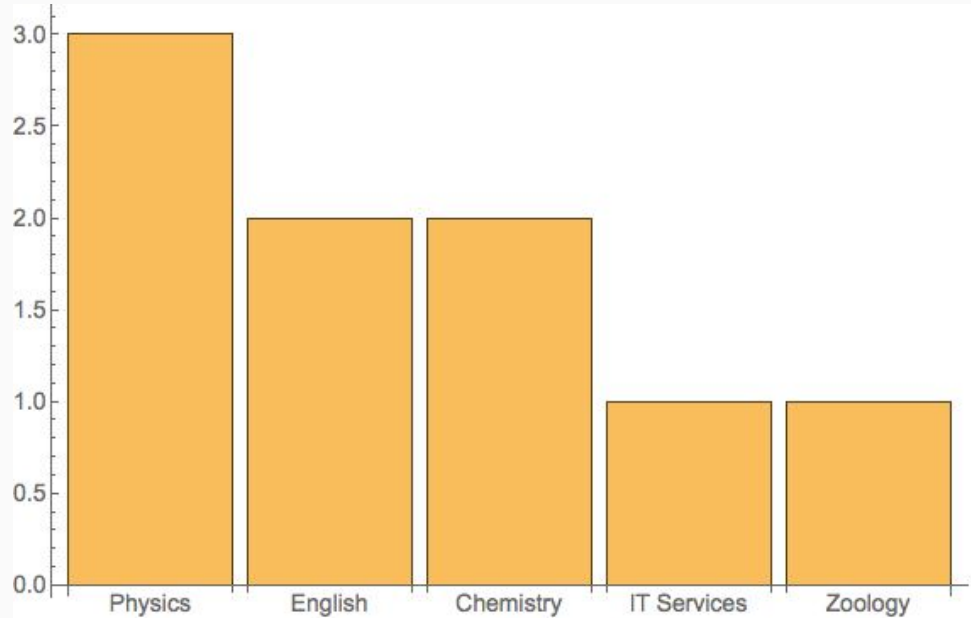
# Graphical Perception Theory



# Graphical Perception Theory

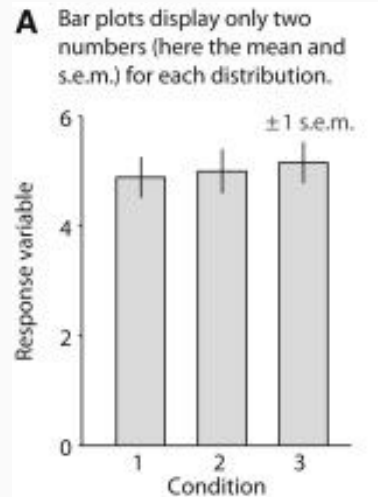


# Graphical Perception Theory



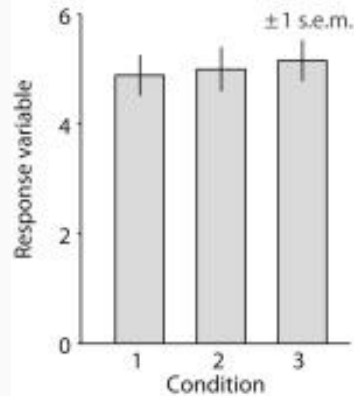
“Pies are evil” is a trope - it’s worth reading [\*Arcs, Angles, or Areas: Individual Data Encodings in Pie and Donut Charts\*](#) if you’re interested

# Graphical Perception Theory

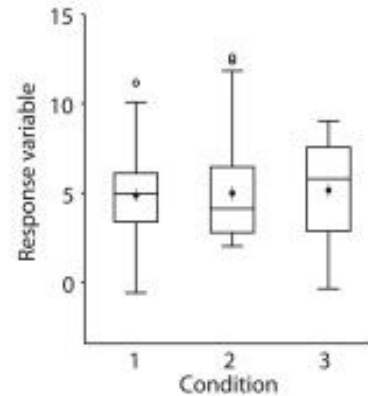


# Graphical Perception Theory

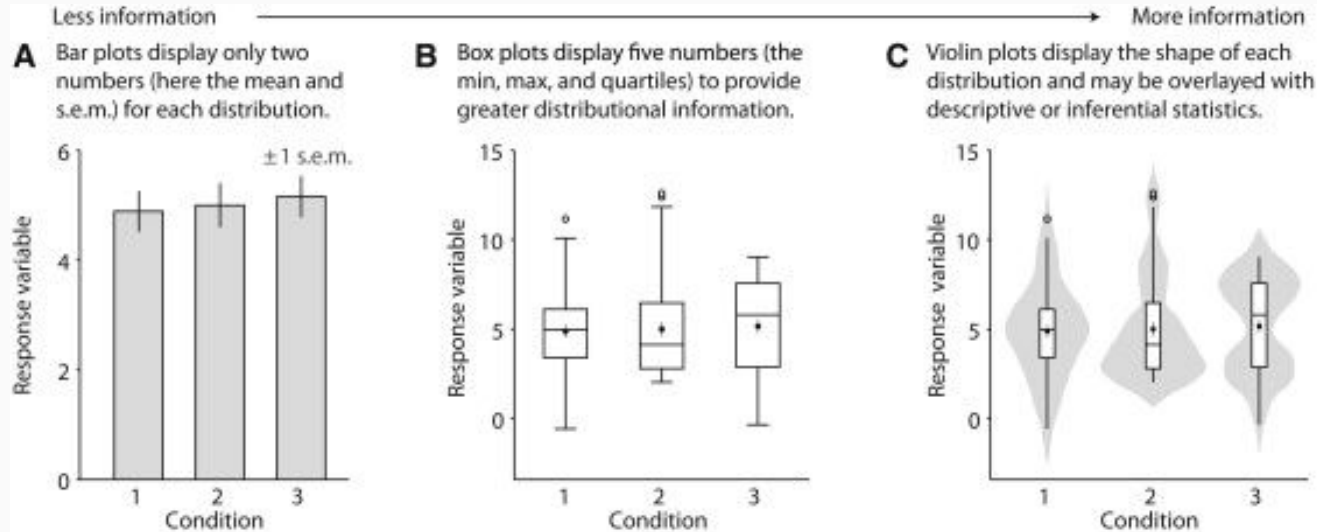
**A** Bar plots display only two numbers (here the mean and s.e.m.) for each distribution.



**B** Box plots display five numbers (the min, max, and quartiles) to provide greater distributional information.

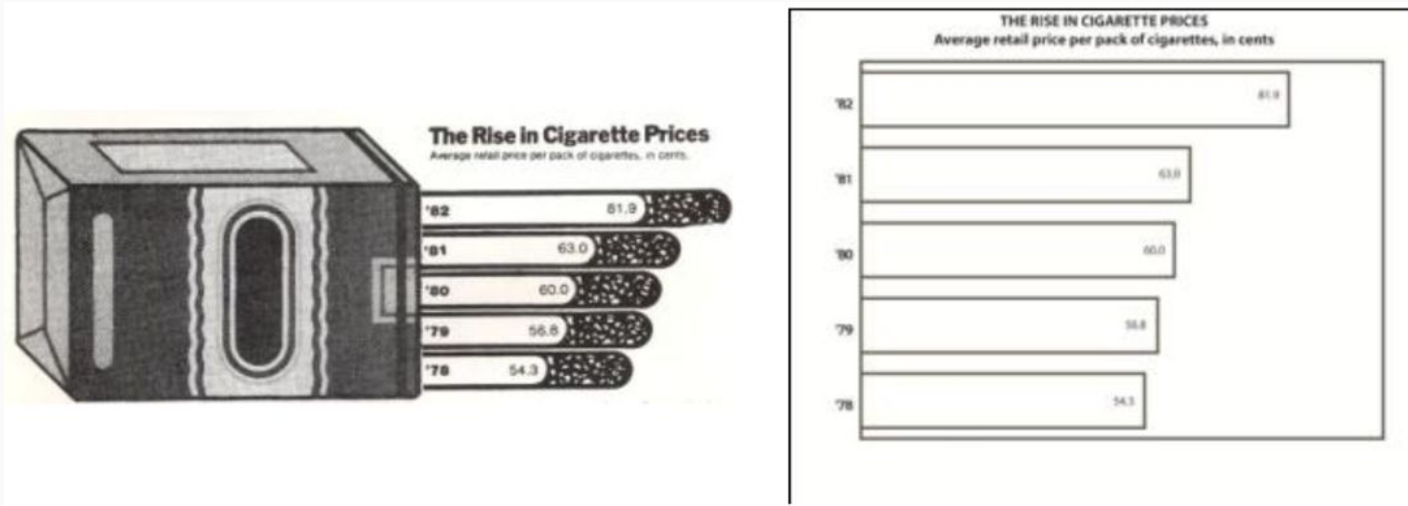


# Graphical Perception Theory



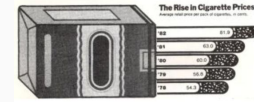
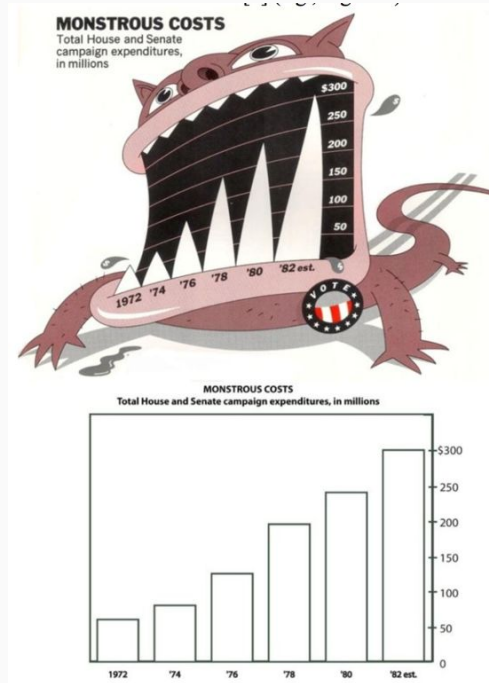


# Chart Junk?



"Kill the frills and get to the point!" [Edward Tufte, 1997]

# Chart Junk?

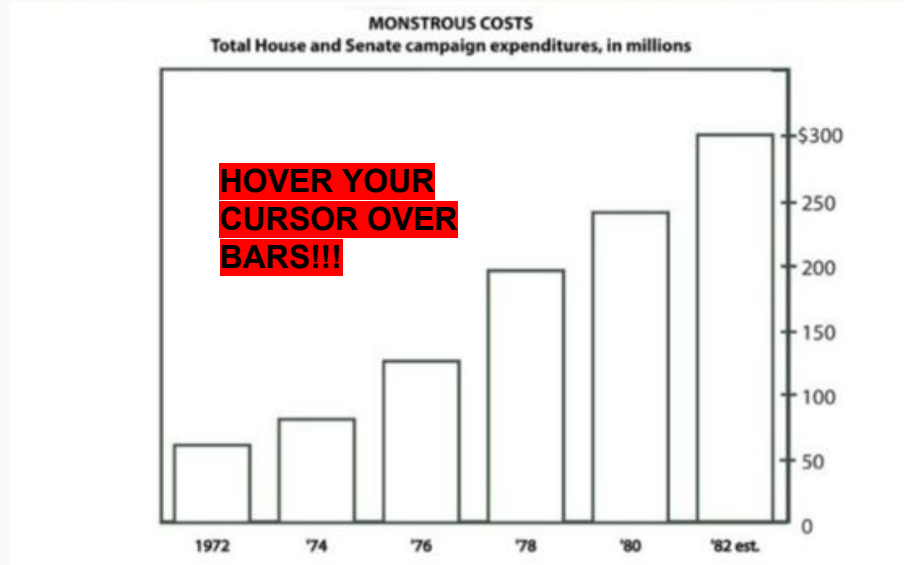


"Kill the frills and get to the point!" [Tufte, 1997]

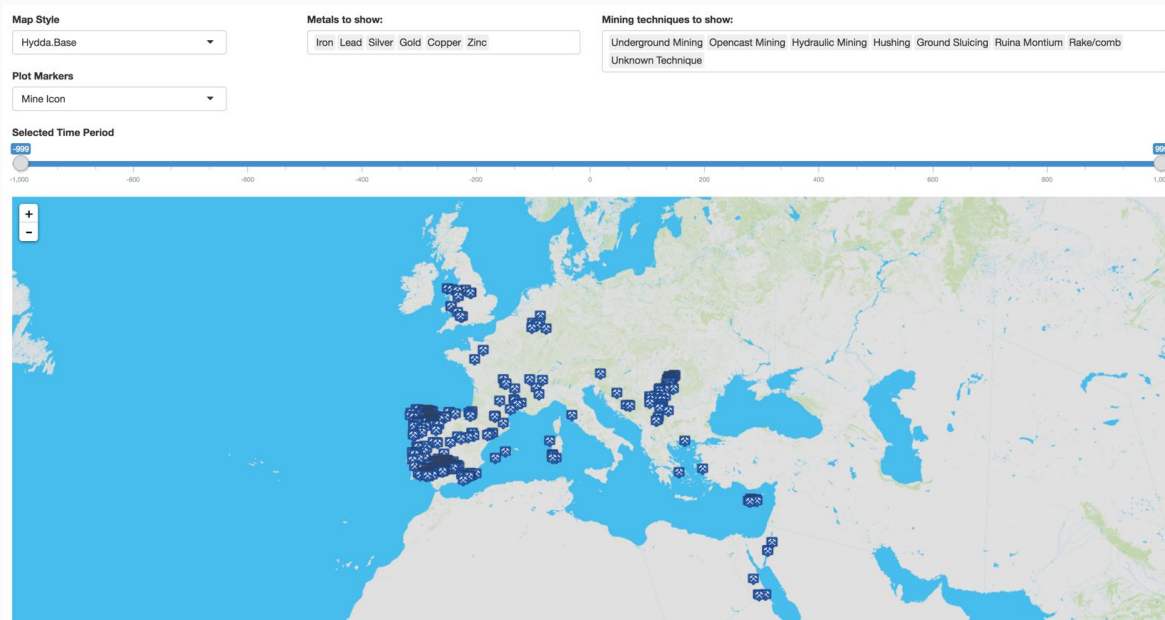
# Chart Junk?



# Making interactivity obvious



# Making interactivity obvious



# Designing with flexibility



**Scott Murray**  
@alignedleft



 Follow

Chart logic reflects a generous assumption by @nytimes that it wouldn't ever get \*this\* dire for Trum

**Hillary Clinton** has an **88% chance**  
of winning the presidency.

Clinton

Trump

2016 Election Forecast »

Updated Oct. 12, 2016

# Designing with flexibility



**Scott Murray**

@alignedleft



Follow

Chart logic reflects a generous assumption by @nytimes that it wouldn't ever get \*this\* dire for Trum

**Hillary Clinton** has an **88% chance** of winning the presidency.

Clinton

Trump

2016 Election Forecast »

Updated Oct. 12, 2016

## Chance of winning



Hillary Clinton

**86.2%**

Donald Trump

**13.8%**



FiveThirtyEight

<http://projects.fivethirtyeight.com/2016-election-forecast>

# Reproducible Research & Visualisations

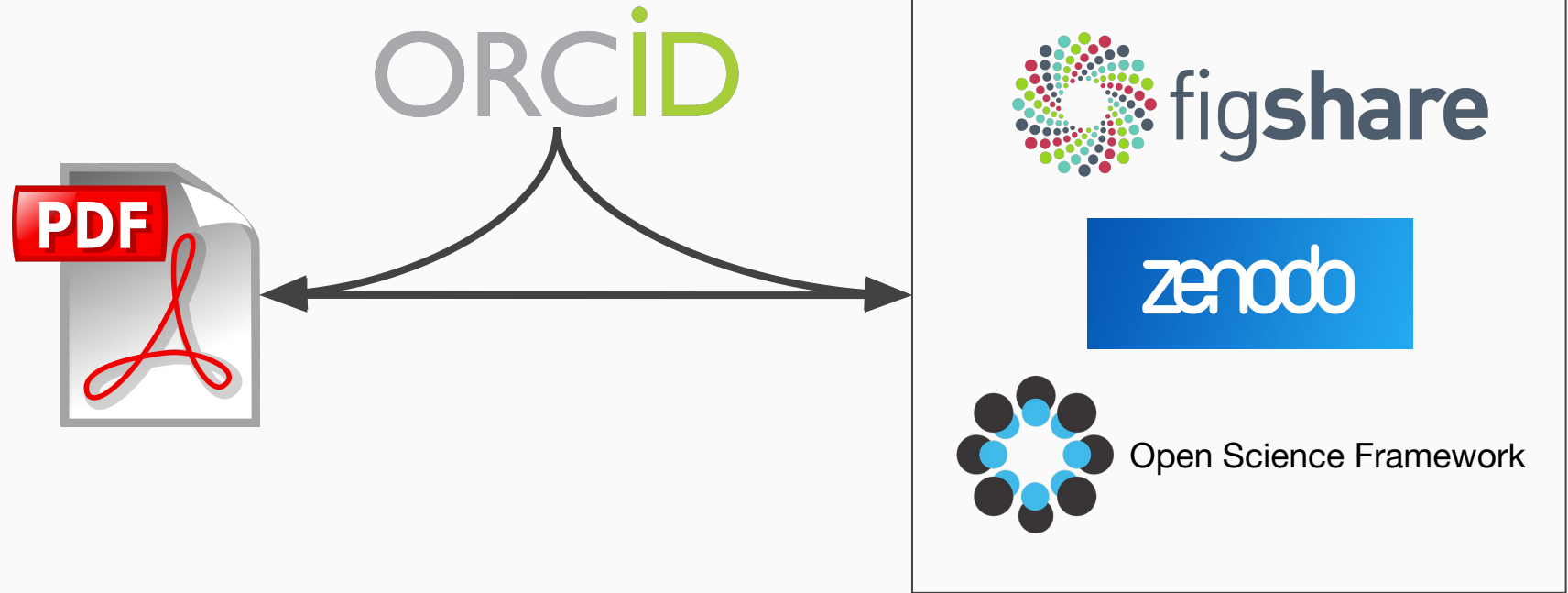
Integrating web-based visualisations into your workflows



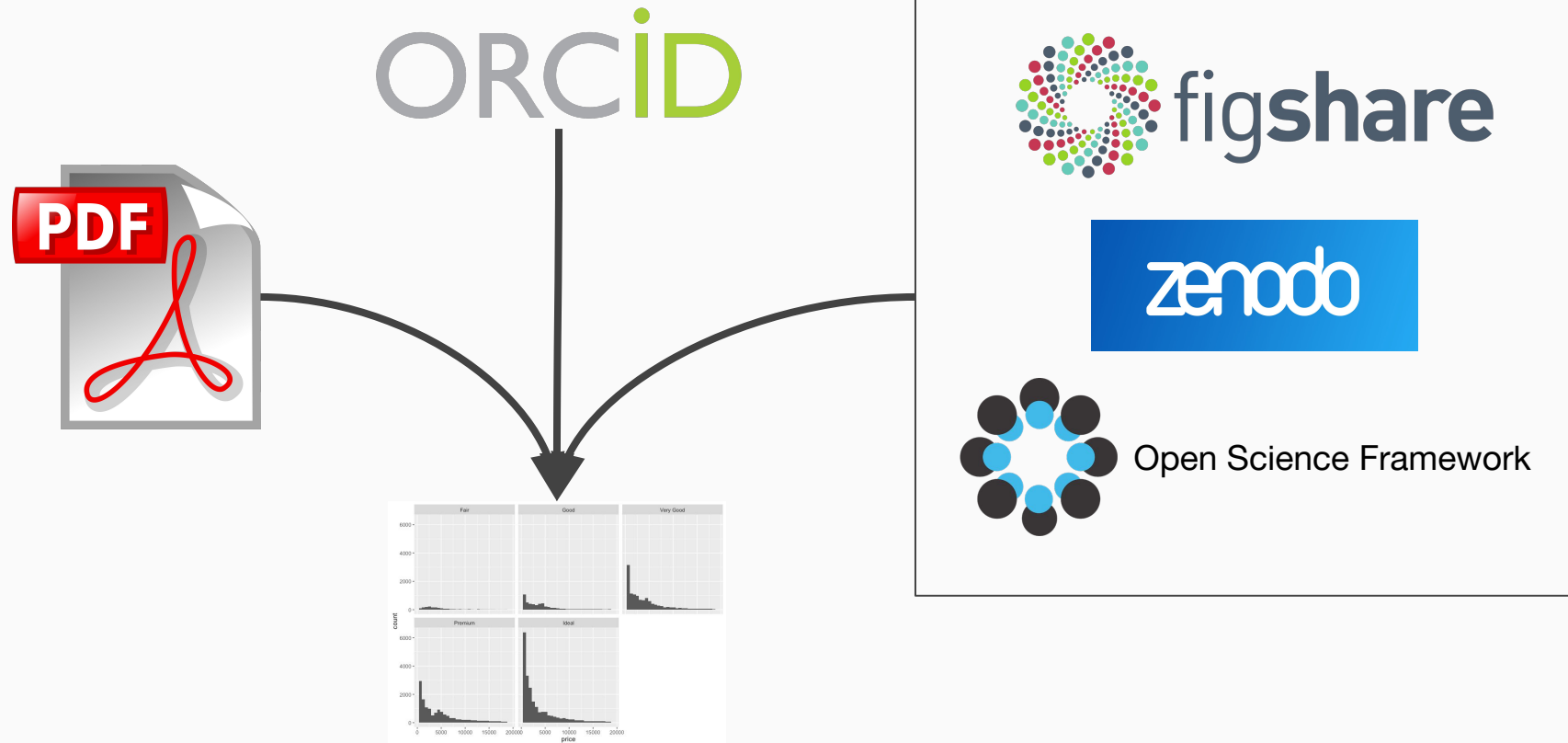
@martijnhadley



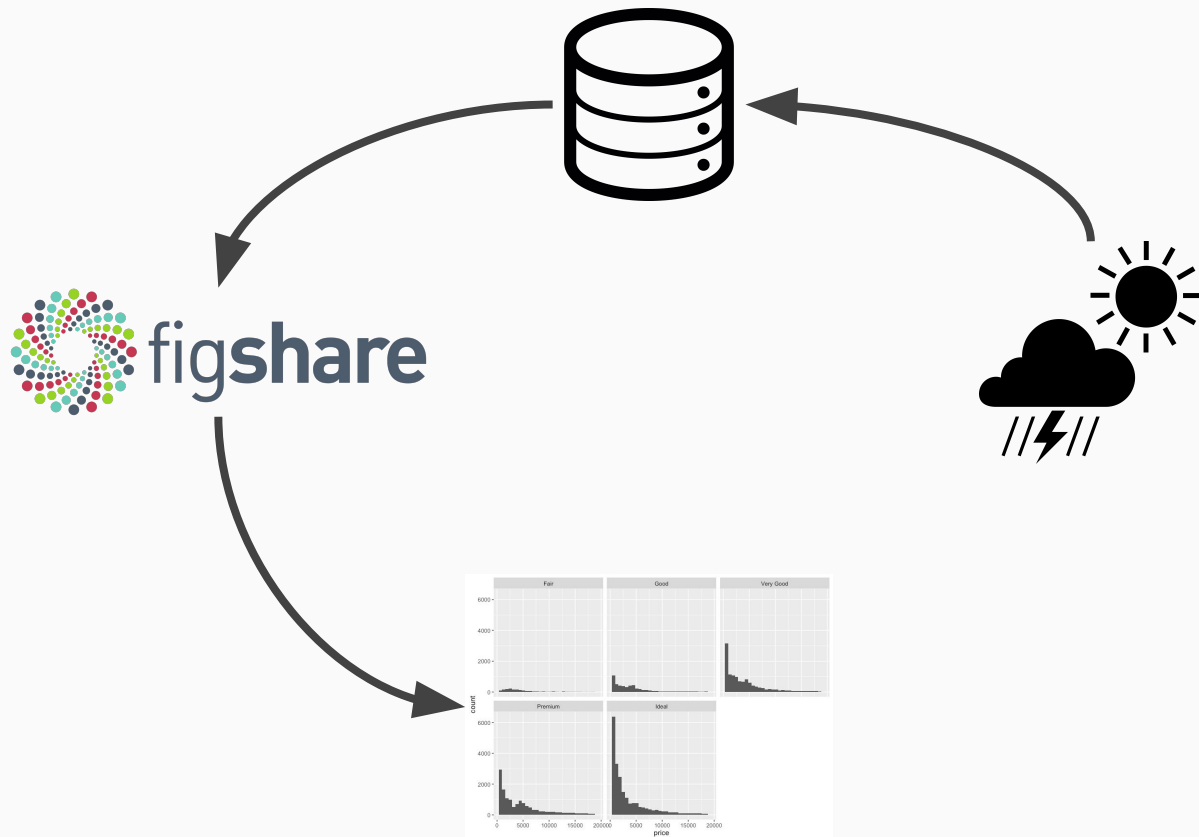
# The citation pyramid



# The citation pyramid



# Living datasets



# Selecting visualisation tools

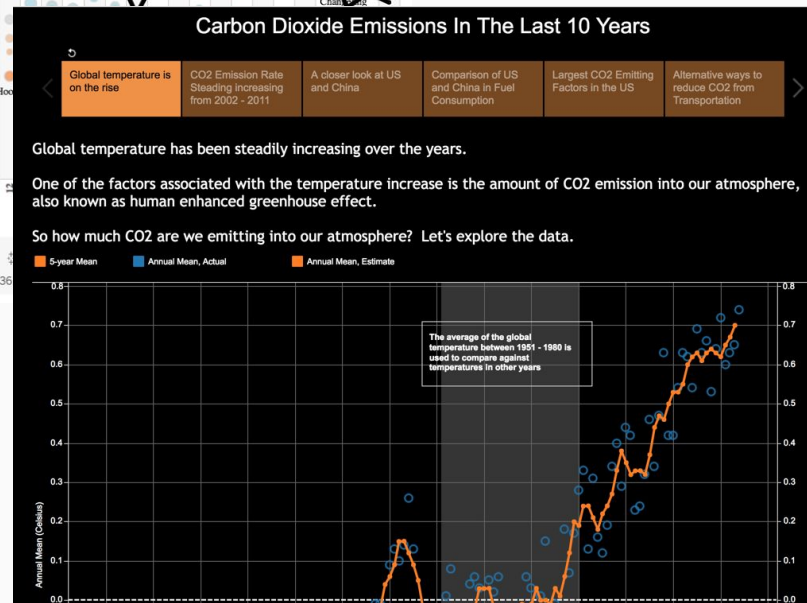
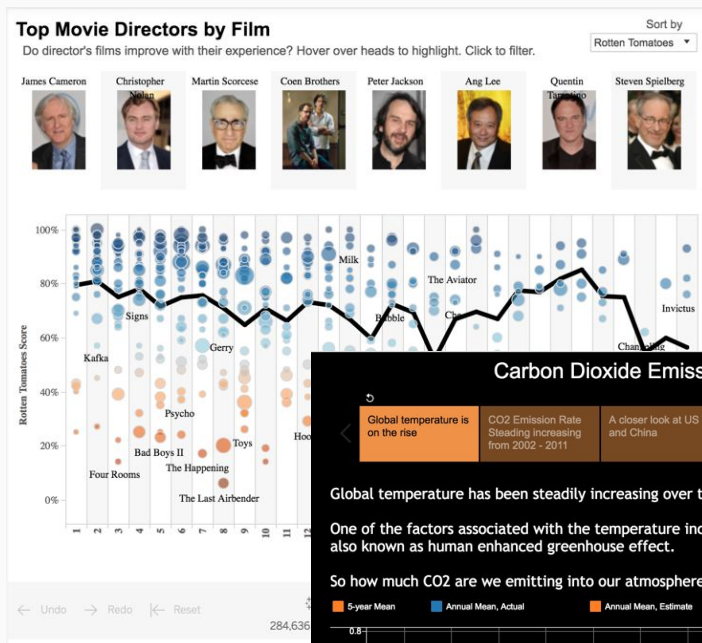
Point & click or script - choose your poison



# Point and Click Tools



- Tableau Public

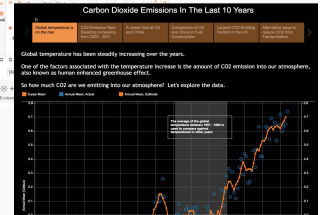
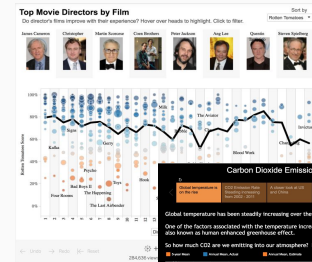


# Point and Click Tools



- Tableau Public

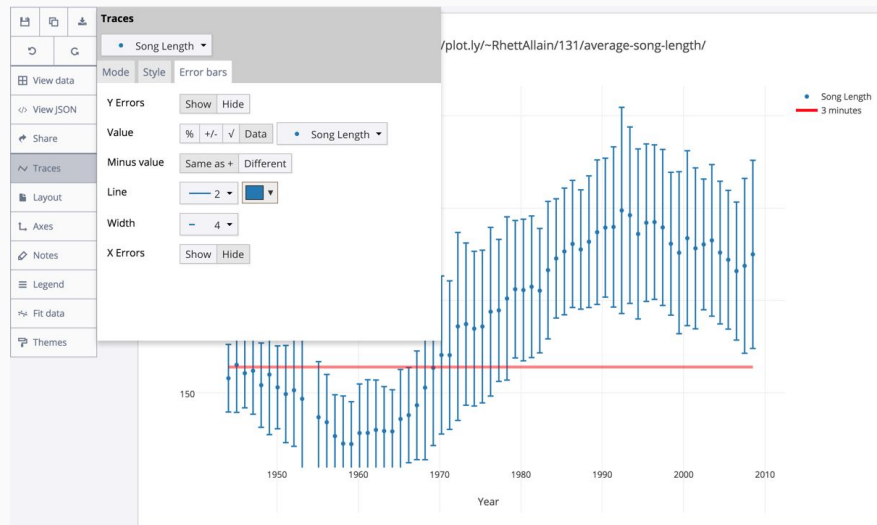
- Tableau is incredibly simple and powerful to use
- All data within visualisations can be downloaded and reused
- Licensing is permissive and does *not* require attribution



# Point and Click Tools



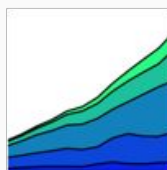
- Tableau Public
- Plotly



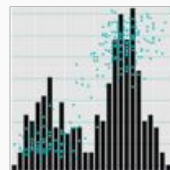
Dashboards



Line and Scatter  
Plots



Filled Area Plots



Graphing Multiple  
Chart Types



Bubble Maps



Lines on Maps

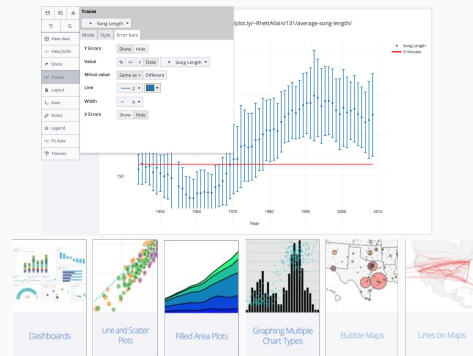
# Point and Click Tools



- Tableau Public
- Plotly



- Public visualisations (and datasets) can be made freely
- Data and visualisation provenance integral to every chart (citable visualisations!)
- APIs for R, Python and more



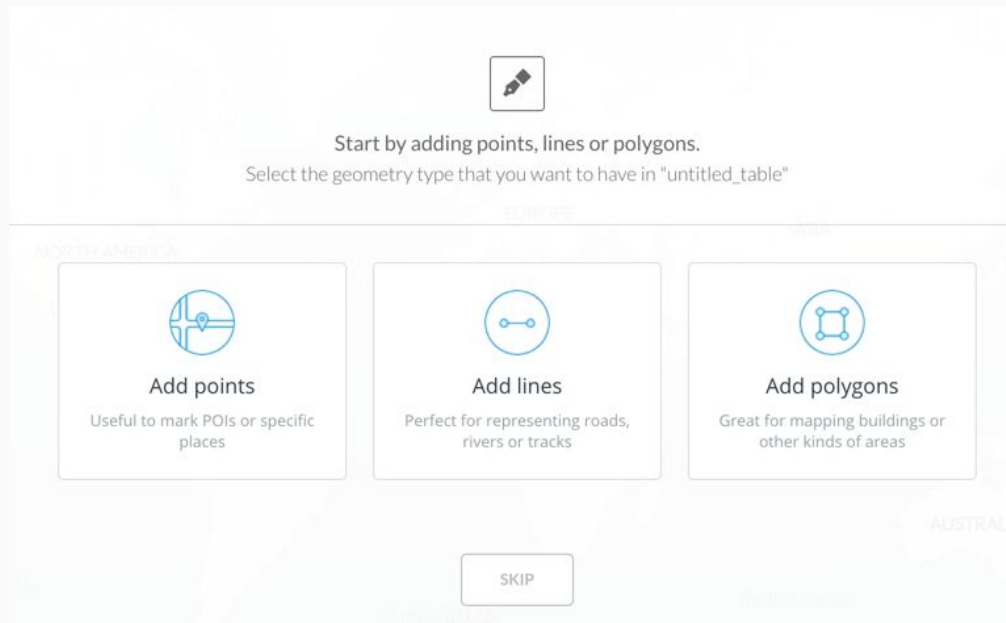


# Point and Click Tools



- Tableau Public
- Plotly
- Carto.com

# CARTO



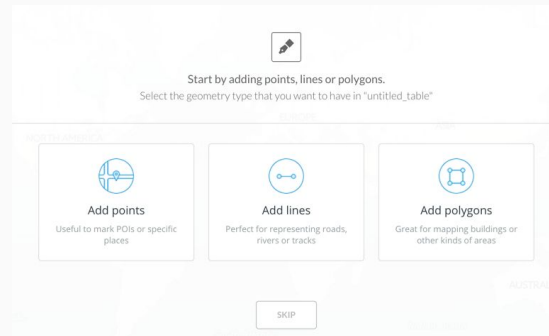
# Point and Click Tools



- Tableau Public
- Plotly
- Carto.com

# CARTO

- Public visualisations (and datasets) can be made freely
- Data can be downloaded without attribution
- More powerful than Google Map Maker, even with a free account



# Point and Click Tools



- Tableau Public
- Plotly
- Carto.com
- ... so many more

- More and more free-to-use point & click services are becoming available
- Long term viability of these services isn't well known - but appear profitable
- Data on these services may be public but shouldn't be considered "open"

# Scripting Languages

- Python



- Bokeh provides a framework for building static and interactive charts
- Interactive charts rely on a technology stack you're responsible for yourself



Bokeh



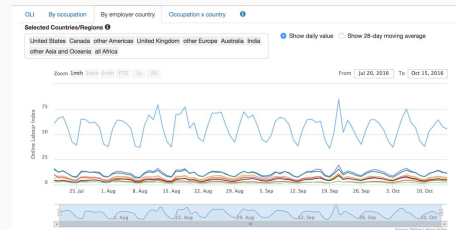
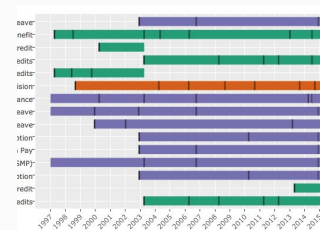
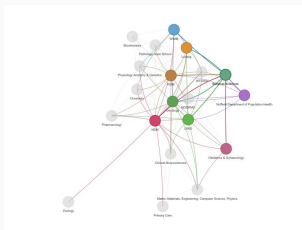
Flask

web development,  
one drop at a time

# Scripting Languages

- Python
- R

- Shiny is a self-contained framework for interactive visualisations built with R
- Building these visualisations requires no knowledge of HTML or JavaScript

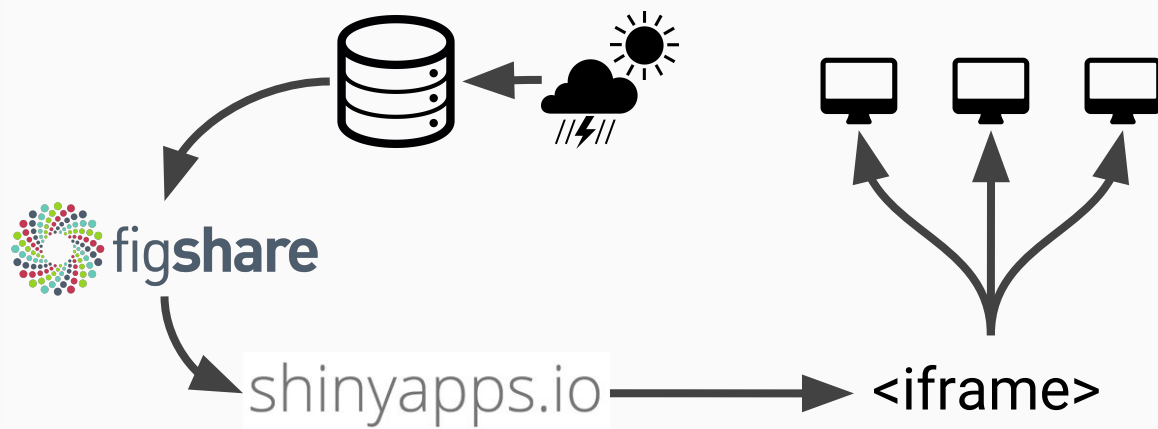


# Scripting Languages

- Python
- R

## shinyapps.io by RStudio

- Shinyapps.io provides hosting for shiny apps
- Free tiers are available, charge is decided by the popularity of shiny apps
- No data is required to live on shinyapps.io



# Data visualisation is important

Data should only be considered visible when it is easily accessible and understood

Thank you for listening

Email:

[martin.hadley@it.ox.ac.uk](mailto:martin.hadley@it.ox.ac.uk)



[@martinjnhadley](https://twitter.com/martinjnhadley)

Slides after here are collections of notes



# Visualisation Blogs and Resources

@agereyes

<https://eagereyes.org/>

 **FLOWINGDATA**

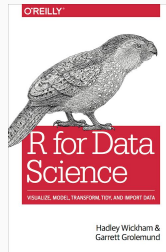
<http://flowingdata.com/>

 **FiveThirtyEight**

<http://fivethirtyeight.com/>



<http://www.visualisingdata.com/>



<http://r4ds.had.co.nz/>

htmlwidgets Tutorial

[http://ox-it.github.io/OxfordIDN\\_htmlwidgets/](http://ox-it.github.io/OxfordIDN_htmlwidgets/)

**Radiant**

<https://radiant-rstats.github.io/docs/>

**Lynda.com**  
FROM LINKEDIN

<https://www.lynda.com/RStudio-tutorials/Creating-Interactive-Presentations-Shiny-R/452087-2.html>\*

\*shameless self-promotion

# GIFS as data visualisations

Animation can get let data speak for themselves

